Junius, Karsten

Working Paper

Limits to industrial agglomeration

Kiel Working Papers, No. 762

Provided in cooperation with:
Institut für Weltwirtschaft (IfW)

Suggested citation: Junius, Karsten (1996) : Limits to industrial agglomeration, Kiel Working Papers, No. 762, http://hdl.handle.net/10419/46880

Kieler Arbeitspapiere
Kiel Working Papers

Kiel Working Paper No. 762

Limits to Industrial Agglomeration

by

Karsten Junius

August 1996

Institut für Weltwirtschaft an der Universität Kiel
The Kiel Institute of World Economics

ISSN 0342 - 0787

Kiel Institute of World Economics

Department IV

D-24100 Kiel, Germany

Kiel Working Paper No. 762

Limits to Industrial Agglomeration

by

Karsten Junius

697204

August 1996

Limits to Industrial Agglomeration

Contents

Abstract

This paper presents an economic geography model to show the spatial effects of economic integration. While other authors mainly focused on the explanation of cumulative causation effects that lead to complete concentration or absolutely equal dispersion of industries, this paper explains why limits to industrial agglomeration can be observed in reality. It argues that cumulative causation effects can be counterbalanced by further centrifugal forces such as land rents, adverse self-fulfilling expectations and congestion effects. For large scale agglomerations, congestion effects may be the most relevant force that stop a cumulative trend towards complete concentration caused by scale economies and trade costs.

JEL-Classification: F12, R12

Keywords: Economic Geography, Agglomeration, Congestion, Location Theory

Karsten Junius
Kiel Institute of World Economics
Duesternbrooker Weg 120
D-24105 Kiel
Tel.: 0431/8814-497
Fax: 0431/8814-524 or 85853
e-mail: junius@ifw.uni-kiel.de

1. Introduction[1]

Since the beginning of the decade, there has been a strong revival of interest in questions where industries locate, why it comes to industrial concentration and agglomeration and how economic integration of regions may influence the distribution of industries between them. This interest was motivated by a strong trend to form regional trading blocs such as the EU single market, NAFTA, Mercosur etc. A new field in economic theory has evolved building on the work of Krugman and Venables (1990, 1993, 1995), Krugman (1991a, 1991b, 1993) and Venables (1996). These papers combine the tools of modern trade theory and the ideas of traditional location theory to formulate so called economic geography models. The main features of economic geography models are the microeconomic foundation of the centrifugal and centripetal forces in an imperfect competition model, the explicit treatment of trade costs and space, and the modeling of externalities as an endogenous outcome of market forces instead of assuming the more elusive technological externalities. In these models, backward and forward linkages[2] induce a cumulative process that can lead to agglomeration of industries in one region. The two main factors for explaining agglomerations are economies of scale and trade costs. While economies of scale are necessary for an explanation of the advantages of industrial concentration, the level of trade costs determines the strength of these advantages. It

---

1 I would like to thank Erich Gundlach and Anthony Venables for helpful comments on an earlier version.

2 As suggested by Hirschman (1958), backward linkages affect the demand for the output of a firm. They result from good access to consumers. A greater number of consumers supports a greater number of firms. Forward linkages affect the cost side of a firm. They result from good access to suppliers of intermediate inputs. The availability of a greater number of goods increases the utility of firms. On the consumer side, forward linkages result from the availability of a large number of consumer goods, which increases real income of workers.

determines whether a cumulative process leads towards agglomeration or further dispersion of industries.

Hence, the economic geography models provide an intuitive explanation for the existence of agglomerations. However, the existing approaches have one problematic implication. Except for the case of zero trade costs and the case of high trade costs, centripetal forces outweigh centrifugal forces. Although equilibria exist for some parameter combinations that describe a center-subcenter distribution of industries, these are in general unstable.[3] Thus, economic geography models usually predict a tendency towards complete concentration of all industries in one region for low trade costs. This overemphasis of the centripetal forces is a serious drawback. Besides the danger of implying too simplistic policy conclusions, the models apparently fail to describe reality adequately. Industrial structures do not seem to diverge in Europe and the US and center-subcenter distributions of industry and population can be observed frequently. Because neither high trade costs nor zero costs of overcoming economic distance matter for European regions and US states, I conclude that centrifugal forces are inadequately modeled in the existing economic geography models.

In this paper, I discuss a number of further centrifugal forces and their theoretical impact on industrial agglomeration. In section 2, I first present what I call a prototype economic geography model. This model allows for consumer preferences for variety as in Krugman (1991b) and the availability of a large number of specialized inputs as in Venables (1996) and

---

3 Also multi-location models like Krugman (1992, 1993, 1994) and Krugman and Venables (1995b) do not find industry distributions that show asymmetric results with different sizes of the industrial sector in different regions. The models by Fujita and Krugman (1995) and Fujita and Krugman and Mori (1995) are the only examples of models that try to consider a hierarchical distribution of industries, albeit in an approach that is illsuited to take account of current trade patterns.

Krugman and Venables (1993).[4] Both forces induce a cumulative process towards industrial agglomeration. The demand of an immobile constant returns to scale sector constitutes a centrifugal force that works against concentration. In section 3, I solve the prototype model for different parameter values, and show how regional economic integration influences the regional distribution of industries. In section 4, I include congestion effects in the prototype model. I show that the inclusion of congestion effects in an economic geography model leads to the emergence of stable center-subcenter patterns of industry location. Section 5 discusses the impact of land rents and commuting costs. Section 6 analyzes the impact of self-fulfilling expectations. Section 7 summarizes and concludes.

2. The Prototype Model

In this section, I present a prototype economic geography model of industry agglomeration that shows the working of centrifugal and centripetal forces. I assume that there are two regions in the economy, which I call core (c) and periphery (p). I mainly refer to the core region and use the indices c and p only where necessary. With different indices, the same equations hold for the peripheral region. For expositional ease, I also refer to the core as the domestic region and to the periphery as the foreign region, without implying that a region equals necessarily a political unit that is distinct from the other region.

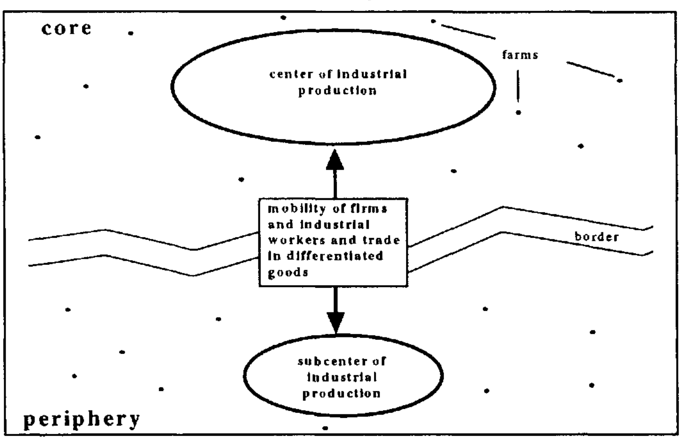
---

4 The model differs from Krugman (1991b) in that it considers the costs and demand linkages that result from the presence of other firms. It differs from Venables (1996) and Krugman and Venables (1993) in that it considers a constant returns to scale sector and from Krugman and Venables (1995a) in that it allows for regional mobility of industrial workers.

Each region has two sectors – an increasing returns to scale sector, which I call the industry, and a constant returns to scale sector, which I call agriculture. The industrial sector consists of firms. The agricultural sector consists of farms. Labor is the only primary factor of production, which consists of workers and farmers. Workers can only be employed by firms. Farmers can only be employed by farms. Thus, there is no mobility between sectors. However, workers are interregionally mobile, while farmers are interregionally immobile.[5]

Initially, the regions are symmetric in all possible aspects except for the size of the industrial sector. The region with the larger industrial sector, i.e. the region with the larger number of firms and workers, is called the core. The size of the arable land is equally distributed, so that the number of farmers living in each region is the same. This means that the core has a higher ratio of firms to farmers. This economy is illustrated in Figure 1.

Figure 1 — The Core-Periphery World



5 This shall take into account that a certain part of a population is immobile and does not or barely responds to economic incentives. The constant returns to scale sector does not necessarily have to be called the agriculture sector, but will so in order to be in line with other economic geography models.

I normalize the total number of workers and farmers in both regions to one so that π is the share of workers (L) in total population:[6]

[2.1] $\pi = L_c + L_p$,

so the share of farmers in total population is $(1-\pi)$. Since each region has the same number of farmers, $(1-\pi)/2$ is the number of farmers per region. For simplicity, I further normalize wages and marginal productivity of farming to one and assume that agricultural output can be transported costlessly within and between regions.

Preferences for the consumption of industrial and agricultural products are described by the standard Dixit and Stiglitz (1977) utility function (U), consisting of a Cobb-Douglas and a CES[7] part:

[2.2] $U = C_M^{\varepsilon} C_A^{1-\varepsilon}$,

and where $C_M = [\sum_{j=1}^{N} c_j^{(\sigma-1)/\sigma}]^{\sigma/(\sigma-1)}$. $1-\varepsilon$ is the share of income spent on the consumption of agricultural products (C_A). ε is the share of income spent on the consumption of industrial goods (C_M), c_j is a single product variety, σ is the elasticity of substitution between the product varieties and N is a large number of potential products.

Production per firm (Q_{cj}) can be described by the input demand function:

[2.3] $Z_{cj} = \alpha + \beta Q_{cj}$, with $\alpha, \beta > 0$,

---

6 See also the appendix for a list of the variables and parameters used in the model.

7 CES means constant elasticity of substitution. In a CES-utility function, the elasticity of substitution is independent of consumer's income, the initial prices of the goods and the number of goods in the economy that have already been consumed. A change in relative prices always leads to a proportional change in the quantities consumed.

where α are fixed costs and β are variable costs. Inputs are a composite of workers and intermediate goods (M). Labor has a share of $1-\mu$ and intermediate inputs have a share of μ in total inputs.

[2.4] $$Z_{cj} = L_{cj}^{1-\mu} M_j^{\mu}.$$

Due to fixed costs, firms produce with internal economies of scale. Industrial goods are differentiated goods. Industrial goods are close, but imperfect substitutes, with σ the elasticity of substitution between the different product varieties.[8] Only some of the infinitely possible product varieties can be produced. In order to save on additional fixed costs, each firm produces only one good. Because no firm wants to share the demand for its good with any other producer, and because goods can be differentiated costlessly, every firm produces a different variety of the industrial good. Consequently, the number of firms equals the number of industrial goods. Firms have some monopoly power, because they are the only supplier of a specific variety. Thus, they face a downward sloping demand curve and set their price with a mark-up over marginal costs. However, free market access guarantees that prices equal average

---

[8] The elasticity of substitution in consumption is defined as the percentage change in the quantity ratio of two goods in consumption induced by a relative price change; i.e. $\left(\frac{p_c}{p_p}\right)^{\sigma} = \frac{c_p}{c_c}$. In the case of a large number of goods, Dixit and Stiglitz (1977) argue that firms neglect their product's marginal contribution on the price index and the total amount of consumers expenditure. The elasticity of substitution then equals the elasticity of demand (e). See also Yang and Heijdra (1993), Dixit and Stiglitz (1993) and d'Aspremont et al. (1996) for a discussion of this assumption. Under monopolistic competition, firms set the price such that marginal revenue equals marginal costs. Consequently, marginal costs are at a positive level. Because firms have some market power, prices are set such that marginal revenue equals $p(1-1/e)$. For a positive marginal revenue, $e>1$ must hold. It follows that under monopolistic competition firms produce on the elastic part of the marginal revenue curve, i.e. where the elasticity of demand exceeds 1. Since the elasticity of demand equals the elasticity of substitution in this framework, $\sigma>1$ has to hold as well.

costs. This leads to the following two conditions for prices of industrial goods, expressed in input units:

[2.5] $\beta + \alpha/Q_{cj} = p_{cj} = (\sigma/\sigma\text{-}1)\beta,\ \sigma > 1.$

The pricing conditions can be used to calculate the output per firm:

[2.6] $Q_{cj} = (\sigma\text{-}1)\alpha/\beta.$

It can be seen that output per firm is independent from the number of domestic customers and from domestically available inputs. It is always the same for all firms in both regions.

In the output equation, the elasticity of substitution also determines the level of scale economies (EOS). The degree of EOS can be measured by the ratio of average costs (AC) to marginal costs (MC). Using equations [2.5] and [2.6], this ratio can be expressed as $\sigma/(\sigma\text{-}1)$, the mark-up that monopolistic firms charge on top of MC. Prices additionally equal AC, since firms earn zero profits in monopolistic competition. The mark-up compensates firms for the fixed costs of production. The higher σ, the smaller is the mark-up. With a small mark-up, firms have to produce a larger output to break-even. Thus, σ is positively correlated with output per firm. The higher the output, the lower is AC/MC. Consequently, a higher σ indicates lower EOS.

The assumption that industrial goods are differentiated instead of homogenous goods leads to the crucial point that makes agglomeration advantageous in economic geography models. It affects firms as well as households, because both have preferences for variety. Firms prefer variety, because the availability of a large number of differentiated inputs makes it possible to use varieties that best fit the production process. This reduces the cost of production.[9]

---

[9] See Ethier (1982) and Markusen (1989), who introduced Dixit-Stiglitz type preferences for variety on the producer side.

Consumers prefer variety, because they like to consume as many different industrial products as possible. Thus, diversified consumption increases both consumers' as well as firms' utility.

To simplify further, I assume symmetry of consumption and intermediate goods. First, the elasticity of substitution in production and consumption is the same. Second, industrial output can be used for consumption or as an intermediate good. Thus, intermediate inputs and consumption goods are defined identically as:

[2.7] $$M = [\sum_{j=1}^{N} m_j^{(\sigma-1)/\sigma}]^{\sigma/(\sigma-1)} = C_M = [\sum_{j=1}^{N} c_j^{(\sigma-1)/\sigma}]^{\sigma/(\sigma-1)},$$

where m_j are single product varieties.[10]

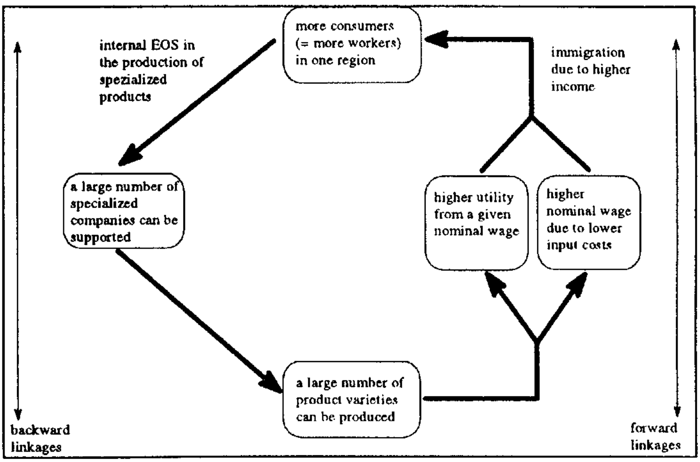
Once there is an unequal regional industry distribution with more workers and firms in the core, demand for goods is higher in the core as well (see Figure 2). Due to fixed costs, no region ever produces all possible and sellable goods. A larger number of workers and thus a higher domestic demand means that a larger number of firms can be supported in the core. A larger number of firms means that more different goods are produced in the core, since all firms produce different varieties. With more goods being produced in the core, less goods have to be imported. Thus, for fewer goods additional transport costs have to be paid. This satisfies firms as well as consumers, because both have preferences for variety, such that they always want to use goods from both regions. Firms can produce more cheaply if they have a lot of different varieties, out of which they can use the most suitable for their specific needs. Due to equation [2.7], this reduces total input demand and raises nominal wages of workers in the core. The outer circle in Figure 2 presents this argument. The inner circle shows the effect of consumer

---

10 Equation [2.7] implies that every consumer always buys a small amount of each available product. This should be interpreted in the way that consumption follows a normal distribution over all goods or that the condition is met on average at least. By symmetry, the same applies for producer preferences.

preferences for variety. A larger number of varieties yields a higher utility and real wages from a given nominal wage. Thus, with more varieties produced in the core, wages are rising and are making it even more attractive. This leads to migration from the periphery to the core, which reinforces the process. As a result such a cumulative process may eventually lead to complete concentration in the core. This completes the centripetal forces of the model.

Figure 2 — Circular Causation Leading to Industrial Agglomeration



The demand from the agricultural sector constitutes the centrifugal force in the model. Firms may wish to locate near peripheral farmers in order to be closer to final consumption and local demand for their products. A firm is facing higher local demand from farmers in the periphery, because the ratio of firms per farmers is higher in the periphery than in the core. Thus, firms might have an incentive to locate in the periphery instead of in the core. In order to evaluate whether centripetal or centrifugal forces dominate some further equations have to be derived.

Firms employ labor and intermediate inputs in such a proportion that the ratio of wages and the price index of intermediate inputs equals the marginal rate of technical substitution of intermediate goods and labor:

[2.8] $$\frac{w_c}{T_c} = \frac{1-\mu}{\mu}\frac{M_j}{L_{cj}}$$

T is the price index for industrial goods. A firm uses intermediate goods from its own region and the other region. The relationship with which industrial goods from the two regions are used depends on the price of the goods and the elasticity of substitution between the goods. The price for domestic varieties equals their marginal costs times the mark-up. For foreign varieties, trade costs have to be paid in addition. Prices for goods from one region are always the same, so that always the same amount of all goods from one region is used. Because the price of foreign goods includes trade costs, the amount of each foreign good used is the smaller, the higher are trade costs. For simplicity, I assume „iceberg-type" trade costs. This means that $\tau > 1$ goods have to be shipped for one good to arrive. Iceberg-type trade costs have the advantage that no further transportation sector has to be added to the model. The price of an imported good, therefore, is τ times the price that it is charged in its region of origin.[11] The regional price index for industrial goods results from the weighted prices of domestic and foreign intermediate goods:

[2.9] $$T_c = [\frac{N_c}{N_c+N_p}(\frac{\sigma}{\sigma-1}MC_c)^{1-\sigma} + \frac{N_c}{N_c+N_p}(\tau\frac{\sigma}{\sigma-1}MC_p)^{1-\sigma}]^{1/(1-\sigma)}.$$

The price index in the core is the lower, the higher is the percentage of domestic goods as a fraction of all goods produced in the economy ($N_c/(N_c+N_p)$), since the price for foreign goods

---

11 Iceberg-type trade costs have first been suggested by Samuelson (1954). However, the underlying idea of real trade costs goes back to von Thünen (1826).

includes the additional transport costs. Thus, the price index for intermediate inputs is always lower in the core.

In the appendix, it is shown that marginal costs depend on the weighted average of wages and the price index for intermediate goods:

[2.10] $$MC_c = w_c^{1-\mu} T_c^{\mu} \beta (1-\mu)^{\mu-1} \mu^{-\mu} .$$

It is also shown that the number of firms and different goods produced in a region can be expressed as:

[2.11] $$N_c = \frac{L_c}{(\alpha\sigma)} [\frac{\mu}{1-\mu} \frac{w_c}{T_c}]^{\mu} .$$

In order to solve for the static general equilibrium of the economy expressions for total income and sectoral wages have to be found. Wages in the agricultural sector equal their constant productivity, which was defined to be one. Income in the agricultural sector per region, therefore, equals the farmers\` share in the total labor force, which is $(1-\pi)/2$ per region. Total income consists of the wage sum in the industrial sector ($w_c L_c$) and income in the agricultural sector:

[2.12] $$Y_c = w_c L_c + (1-\pi)/2 .$$

Total expenditure on industrial goods by consumers and firms in the core (X_c) equals the share ε of the regions income plus the expenditure on intermediate goods ($M_j T_c$). By the use of equation [2.8] expenditure on intermediate goods can be expressed in terms of the sum of wages, such that :

[2.13] $$X_c = \varepsilon Y_c + \frac{\mu}{1-\mu} w_c L_c .$$

In the appendix it is shown that the nominal wage rate is:

[2.14] $$w_c = \frac{(1-\mu)N_c}{L_c(N_c+N_p)}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\left[X_c\left(\frac{T_c}{MC_c}\right)^{\sigma-1} + X_p\left(\frac{T_p}{\tau MC_c}\right)^{\sigma-1}\right].$$

From equations [2.9–2.14], one can determine the price indices, number of firms, income, sales and the wages per region for a given level of the exogenous parameters as well as a given initial distribution of workers. Real wages, ω_c, are calculated as:[12]

[2.15] $$\omega_c = w_c/T_c^{\varepsilon}.$$

In this prototype economic geography model the real wage for workers in the periphery relative to the real wage in the core (ω_p/ω_c) determines whether the core or the periphery is a more attractive place to locate. Workers and firms have an incentive to locate in the core if the ratio is below one. In this case centripetal forces dominate centrifugal forces. If the ratio is above one, workers and firms have an incentive to locate in the periphery. In this case centrifugal forces dominate.[13]

---

[12] Real wages are calculated by dividing nominal wages by the consumer price index. The consumer price index is distinct from the price index of industrial goods, because it additionally includes agricultural goods. For Cobb-Douglas preferences and a price of agricultural goods normalized to one, it can be expressed as $T_c^{\varepsilon}1^{1-\varepsilon}$, which reduces to T_c^{ε}.

[13] In this monopolistic competition type of economy, all possible profits are dissipated by the free entry of firms. The revenue net of the costs of intermediate inputs is distributed to workers through wages. Thus, it seems somewhat meaningless that firms might relocate to the other region when workers move due to regional differences in real wages. However, with zero moving costs and zero costs of setting up the production of goods, the assumption that firms always move in proportion to workers is no restriction for the model. If so the calculated wage can be interpreted as the maximum wage that firms can pay without making losses.

3. The Mechanics of the Prototype Model

In order to evaluate whether centrifugal or centripetal forces dominate, I solve equations [2.9]–[2.15] simultaneously for both regions and various sets of exogenous parameters.[14] From that I calculate the relative real wage of the periphery. This yields a static equilibrium of the model for a specific set of parameters. In this static equilibrium, the relative real wage indicates the direction of the dynamics in the economy. Workers move according to wage differentials and make larger the industrial sector of the region that pays higher wages. This sets in motion a process leading to further dispersion or further agglomeration of firms and workers. Consequently, the relative real wage indicates whether centripetal or centrifugal forces dominate. A ratio of peripheral to core wages below one in the static equilibrium indicates that centrifugal forces outweigh centripetal forces. In this case trade amplifies differences between regions (and does not lead to factor price equalization). As a result, this model can explain where firms locate, independently of technological or taste differences and despite factor and goods mobility.

Centripetal and Centrifugal Forces

I start by analyzing how the strength of different parameters influences the strength of the centrifugal and centripetal forces. This is important in order to fully understand the working of the model, to see how sensible it is to variations of the parameters and how robust its results

---

[14] To solve the model, the Newton procedure on Mathematica for Windows Version 2.2.3 is used. Some analytical explorations of an economic geography model can be found in Krugman (1991b, 1992).

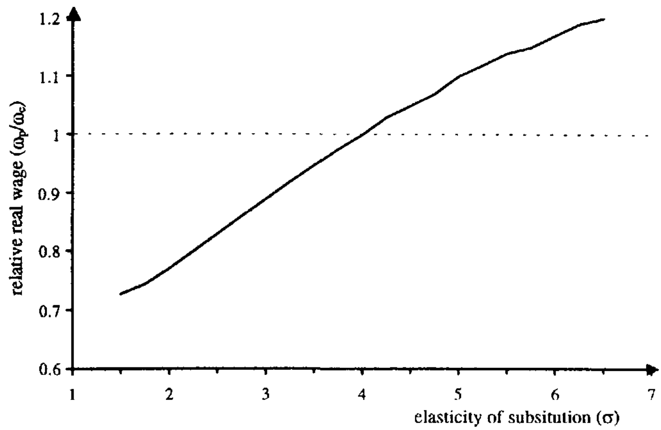
are. The strength of the model is that symmetry between the two regions is assumed for almost all parameters. The size of the industrial sector is the only parameter where the symmetry between the two regions is broken. By the knowledge of how the centripetal and centrifugal forces are influenced by variations of the parameters, one can assess situations where the symmetry is broken at another point.[15]

In Figure 3, I have solved the model for various levels of σ, the elasticity of substitution of the single product varieties. On the horizontal axis, the exogenously predetermined level of σ is shown. On the vertical axis, the endogenously determined level of the relative real wage of the periphery is shown for which the economy is in a static equilibrium. With high elasticities of substitution, workers have low preferences for variety and react very sensitively towards price differences between goods. In the limit of an infinite elasticity of substitution, perfect competition results and consumers buy only the cheapest product variety. Because foreign varieties include trade costs, the amount of domestic varieties bought by consumers is the higher, the higher is the elasticity of substitution. If consumers only buy varieties of their own region, periphery firms have higher revenues than core firms, because of the smaller number of firms per farmers in the periphery. Since this translates into higher wages for workers, relative wages are higher in the periphery, the higher is the elasticity of substitution. Moreover, a

---

[15] Amiti (1995) analyzes such cases. She shows that, if goods differ in their substitutability, the periphery can realize scale economies to a lesser extent than the core, due to a smaller home market. Thus, the periphery would be a net exporter of high elasticity goods and a net importer of low elasticity goods. In addition, she shows that the periphery would specialize in labor intensive goods, if industries differ with respect to factor intensities. Furthermore, the periphery would specialize in the production of goods that are subject to lower transport costs, if industries differ with respect to transport costs.

Figure 3 —Relative Wage for Alternative Levels of Scale Economies, Indicated by the Elasticity of Subsitution

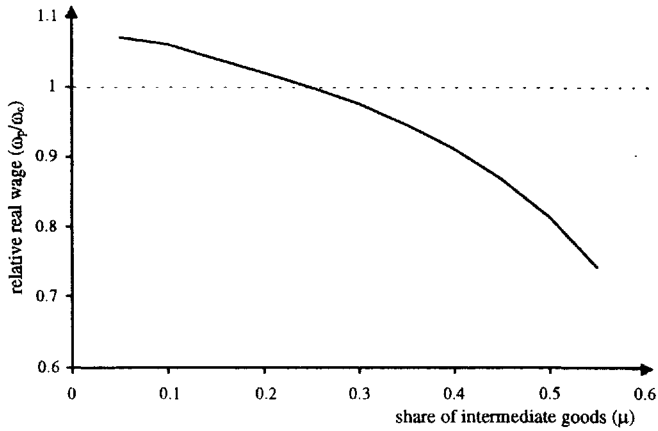


($\alpha = 0.1$, $\beta = 0.4$, $\varepsilon = 0.4$ $L_c = 0.45$, $L_p = 0.25$, $\mu = 0.25$, $\tau = 3$, $\pi = 0.7$)

higher elasticity of substitution means lower economies of scale. With high economies of scale, the division of labor is more advantageous and a region benefits from a large number of firms, of which each is specialized in the production of one good. Since there is a lower number of firms in the periphery, being in the core is the less attractive, if economies of scale do not matter for the production of goods that much. Thus, Figure 3 shows that the relative wage of the periphery is the higher, the lower are economies of scale, i.e. the higher is the elasticity of substitution.

Figure 4 shows the impact of alternative shares of intermediate inputs in the production of goods. A higher μ means that firms use relatively more intermediate goods and less labor. Thus, it shows the importance of forward linkages. Forward linkages are stronger in the core, where more domestically produced intermediate goods are available. Consequently, Figure 4 shows that higher values of μ are associated with a lower relative wage of the periphery.

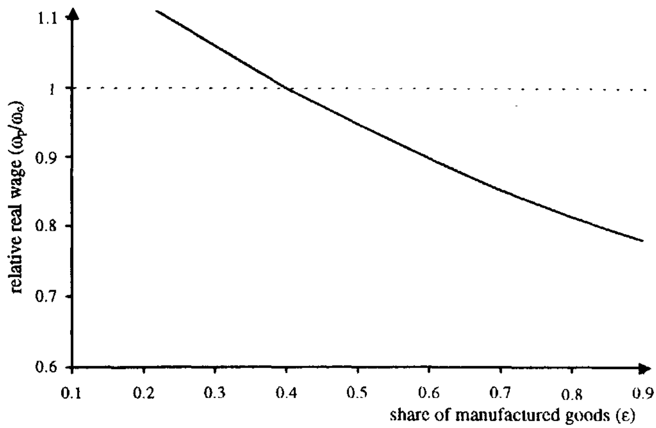
Figure 4 —Relative Wage for Alternative Shares of Intermediate Goods in Total Inputs



($\alpha = 0.1$, $\beta = 0.3$, $\varepsilon = 0.4$, $L_c = 0.45$, $L_p = 0.25$, $\sigma = 4$, $\tau = 3$, $\pi = 0.7$)

Figure 5 shows the impact of alternative shares of industrial goods in total consumption. The higher the share, the lower the relative wage of the periphery. Essentially the same argument as with the importance of intermediate goods applies to the importance of consumer goods. Since the periphery produces less varieties domestically than does the core, it also has to import a larger share of consumption goods. However, imported goods are subject to trade costs. Consequently, the higher the share of industrial goods in consumption, the higher the share of imported goods and the higher the consumer price index. This in turn lowers the relative wage of the periphery.

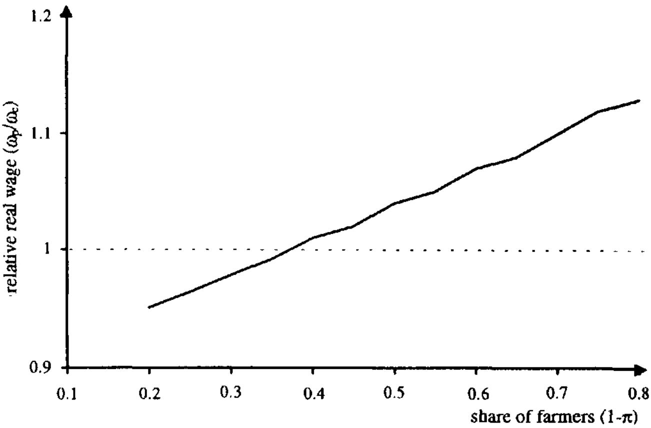
Figure 5 —Relative Wage for Alternative Shares of Manufactured Goods in Total Consumption



(α =0.1, β = 0.3, L_c = 0.45, L_p = 0.25, μ = 0.25, π = 0.7, σ = 4, τ = 3)

Figure 6 shows the relative wage for alternative shares of farmers in the economy. Firms may choose to locate close to peripheral farmers, in order to be close to local consumer demand. Being close to local demand allows firms to avoid trade costs when selling their goods. The assumption that there are fewer firms per farmers in the periphery means that firms face higher demand per firm from local farmers. Thus, the demand of farmers is the centrifugal force of the model that explains, why not always all production takes place concentrated in the core. This force is the stronger, the more dispersed local demand is; i.e. the larger the share of farmers in the economy, the more important for firms is farmers' demand for final goods, and the less important is demand from workers and other firms. Thus, Figure 6 shows that the higher the share of farmers in the total labor force, the higher the relative wage of the periphery.

Figure 6 —Relative Wage for Alternative Shares of Farmers in the Total Labor Force

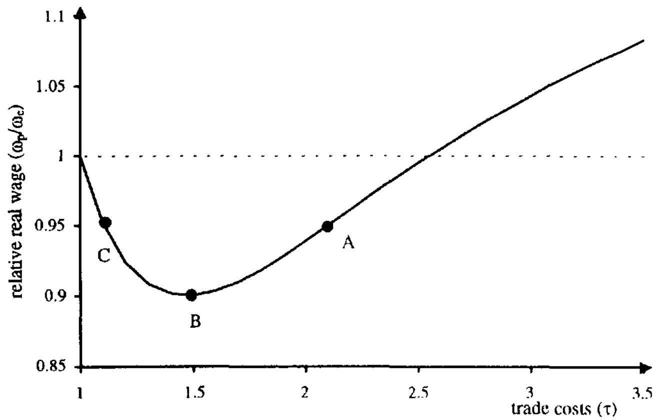


($\alpha = 0.1$, $\beta = 0.4$, $L_c = \pi \cdot 45/70$, $L_p = \pi \cdot 25/70$, $\mu = 0.2$, $\sigma = 5$, $\tau = 2.2$, $\varepsilon = 0.75\pi$)

The Impact of Trade Costs

Trade between the two regions is subject to tariffs, transport costs, and other costs of overcoming economic distance. I subsume all of these under trade costs. The level of trade costs determines whether centrifugal or centripetal forces dominate - and, consequently, determines where industries locate or where they relocate to. In Figure 7 one can see that for zero trade costs, which means $\tau = 1$, wages are the same in both regions. This arises from the fact that location does not matter if nobody has to pay extra trade costs for importing intermediate or final goods. It does not matter where a firm produces its goods and whether it is close to other firms and consumers, because additionally technologies and effective units of labor were assumed to be the same in all regions and technological spillovers were not considered. With relative wages equal to one, nobody has an incentive to move and the distribution of industries is determined by the initial distribution of workers.

Figure 7 — Relative wage for alternative levels of trade costs



($\alpha = 0.1$, $\beta = 0.2$, $L_c = 0.45$, $L_p = 0.25$, $\mu = 0.4$, $\sigma = 6$, $\pi = 0.7$, $\varepsilon = 0.5$)

For high trade costs, few trade takes place. Being in the periphery then means facing relatively less competition and having better access to local farmers' demand. Goods from the core, which include trade costs, are more expensive than goods produced in the periphery. Peripheral farmers mainly demand goods from peripheral firms. Because there are less firms in the periphery, each producer faces a larger proportion of peripheral farmers' demand for industrial goods than core firms face from core farmers. This translates into higher revenues of peripheral firms in the case of high trade costs. Because all possible profits are distributed through wages the peripheral region is able to pay higher wages than the core.[16]

[16] Note that production costs remain higher in the periphery. This results from a higher price index for consumer goods as well as intermediate inputs. Thus, there is a tension between higher production costs and a higher consumer price index on the one hand and better access to local farmers on the other hand.

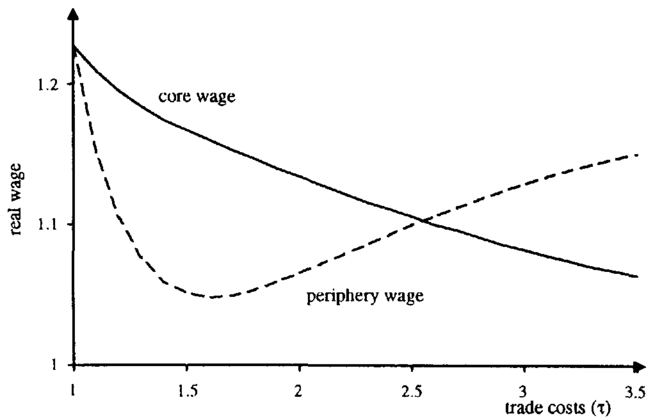
If trade costs fall from a high to a lower level, core firms gain better access to the small market in the periphery and peripheral firms gain better access to the large market in the core. However, this does not improve the position of peripheral firms. Peripheral firms now face more competition from a large number of core firms that each exports a small share of their production to the periphery. This reduces sales of peripheral firms in their own market. In order to match this loss in sales they have to export to the core. However, they have to export a larger share of their production than core firms do, because a few peripheral firms have to match the sales of a large number of core firms. Thus, also a larger share of peripheral firms' output is subject to trade costs. Reducing trade barriers from a high level reduces sales of peripheral firms in their own market and makes a higher percentage of sales subject to trade costs than in the core. Trade costs consume a larger share of revenues from peripheral firms than from core firms, which reduces relative wages in the periphery. If the total reduction of trade costs is sufficiently high, this leads to a situation where the core is able to pay higher wages. This case happens for values of τ for which the curve in Figure 7 is below the dotted line. Centripetal forces then outweigh centrifugal forces and agglomeration results from an adjustment process. This effect is the stronger, the more one goes down the U-curve to point B, which is arbitrarily determined by the exogenous parameters of the model. From then on, the demand of peripheral farmers is becoming less important due to low trade costs. Peripheral farmers then can cheaply be served from the core. Peripheral firms then care more about the access to the large core market. A further reduction of trade costs improves the relative situation of peripheral firms, because it further reduces regional differences in the economic variables like input prices or the consumer price index. With zero trade costs, no costs differences between the two regions prevail so that the relative wage converges to one. Thus, in an early integration process the competition effect dominates and reduces relative wages in the periphery whereas with deeper economic integration this effect peters out and access to the core-market becomes more important. As a result, economic conditions as well as wages converge.[17]

Figure 8 shows the real wages of the periphery and the core instead of the relative real wage as before, pointing to the absolute impact of trade liberalization on real wages. The core region benefits continuously from trade liberalization, because it only reduces costs from trade and the competition of peripheral firms is too small to affect wages in the core negatively. Thus, in relative terms, trade liberalization may negatively affect core wages compared to the periphery for low trade levels. In absolute terms, it is always positive for the core. In the periphery one again finds a U-shaped curve, reflecting absolute losses from trade in an early integration process. However, absolute real wages start rising after some low level of trade costs and approach values that are clearly higher than for the case of high trade costs. They further approach core wages the more trade costs go down.

---

[17] This is a tentative explanation why in the early globalization process – like in the 1970s-public debate mainly focused on the harm that decreasing trade barriers might do to LDCs. Nowadays the debate is reverse. Industrial countries fear the competition from low-wage LDCs. Figure 7 can give an explanation why the attitude towards market integration changed. Lets assume that the world is at point A in Figure 7 and consists out of industrial countries and LDCs only. A reduction of trade costs to point B leads to a fall in relative wages in the periphery, which are the LDCs in this example. Further reduction of trade costs to point C now hurts the core, which are the industrial countries in this example. They still have an absolute advantage in producing industrial goods and LDC workers want to move there, but they cannot justify the same wage premiums as before. Consequently, if a fall in trade costs from point B to C is not accompanied by a fall in the relative wage of industrial countries, they would loose industries to LDCs. In the model, this happens without any change in technologies or productivity and despite the industrial countries absolute advantage in production.

Figure 8 — Real Wages of the Periphery and the Core

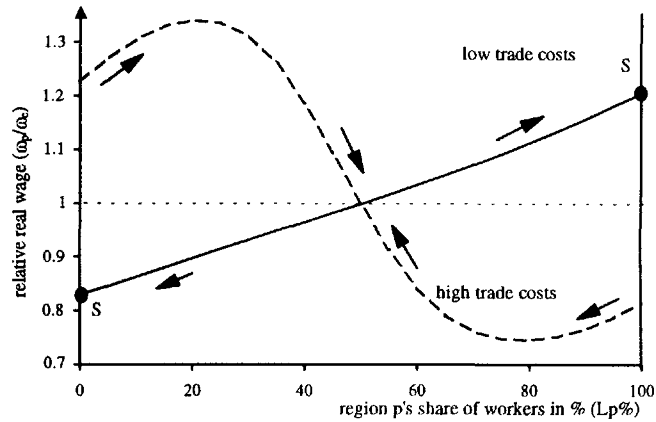


($\alpha = 0.1$, $\beta = 0.2$, $L_c = 0.45$, $L_p = 0.25$, $\mu = 0.4$, $\sigma = 6$, $\pi = 0.7$, $\varepsilon = 0.5$)

The next figures show the relative wage for alternative industry distributions at different levels of trade costs. The horizontal axis exhibits the periphery's share of industry employment. The regions are of same size for an industry share of 50%. Below that the periphery is smaller, above that the periphery is larger. For low trade costs (Figure 9) the larger region always pays a higher real wage. Workers have an incentive to move from the smaller to the larger region, such that two stable equilibria exist, namely those with complete concentration of all firms in either the core or the periphery. A distribution with an equal share of firms in both regions is an unstable equilibrium.

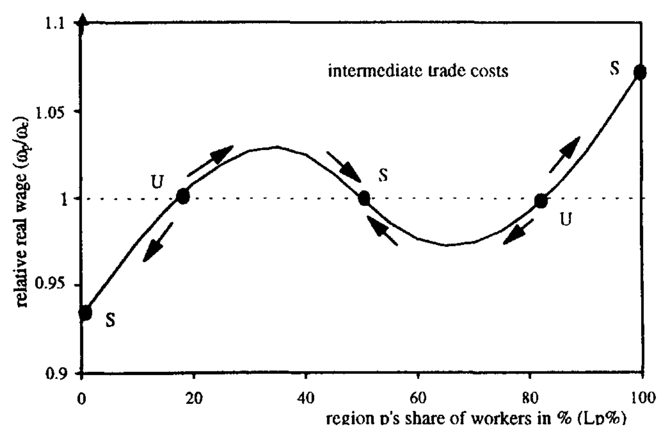
For very high trade costs, Figure 9 shows a reverse situation. For reasons discussed above with high trade costs, the smaller region always pays the higher wage. This yields an incentive for workers and firms to move from the larger to the smaller region until they are both of equal size. It follows that equal distribution of firms is a stable and the only equilibrium of the model.

Figure 9 —Relative Wage for Alternative Industry Distributions (Low and High Trade Costs)



($\alpha = 0.1$, $\beta = 0.2$, $\tau_{low} = 1.5$, $\tau_{high} = 4.5$, $\mu = 0.25$, $\sigma = 5$, $\varepsilon = 0.35$, $\pi = 0.7$)

Figure 10 — Relative Wage for Alternative Industry Distributions (Intermediate Trade Costs)



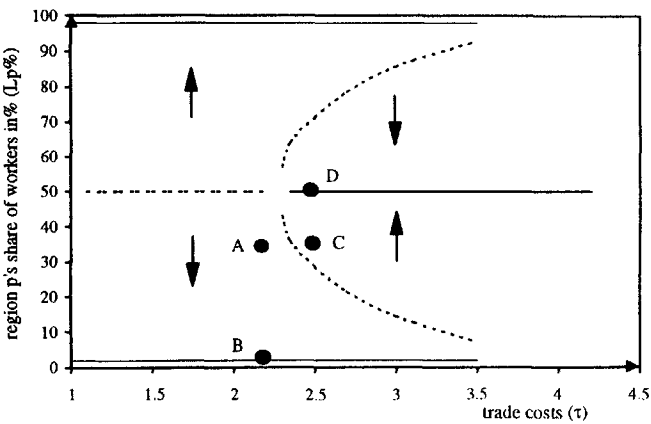
($\alpha = 0.1$, $\beta = 0.2$, $\tau_{int} = 2.2$, $\mu = 0.25$, $\sigma = 5$, $\varepsilon = 0.35$, $\pi = 0.7$)

For intermediate trade costs, Figure 10 shows five possible equilibria. An S indicates a stable equilibrium, an U indicates an unstable equilibrium. As in the case of high trade costs, equal distribution is a stable equilibrium. Workers and firms move to the smaller region until the two regions are of equal size, if they are of somewhat equal size initially. Then the better market access to local farmers in the smaller region outweighs the better forward and backward linkages in the larger region. However, for very unequal initial distributions of firms, the linkage effects are so much stronger in the larger region that the better access to farmers in the periphery cannot match the disadvantage from being far from other producers. This leads to three stable equilibria. Equal distribution as well as complete concentration of firms in one region can result from a migration process, depending on the initial distribution of industries. Between these three stable equilibria there are two unstable equilibria with an unequal distribution of firms between the regions.

The existence of stable and unstable equilibria can also be illustrated in a bifurcation diagram (Figure 11). It shows the equilibrium size of the core for alternative levels of τ. Solid lines represent stable equilibria, dotted lines represent unstable equilibria. For high levels of trade costs equal dispersion of industries is the only stable equilibrium. For intermediate levels of trade costs multiple stable equilibria are possible. Agglomeration results, if the economy starts out very unequally distributed, while equal dispersion results, if the regions are of similar size.

This case illustrates the importance of the critical mass of firms for the development of regions and why a „big-push" can bring about a cumulative process towards an equilibrium with a larger percentage of industrial workers that would not be possible with slow economic development of a region.[18] An industrial sector will only be able to survive in a region, if enough other firms start producing at the same time or the immobile demand is very high. Without the backward and forward linkages provided by the existence of other firms, the development of an industrial sector is not sustainable as maximum wages that a firm can pay are lower than in the larger region at intermediate levels of trade costs. This means that a continuous or slow development of an industrial base cannot be successful for intermediate levels of trade costs. Only if the percentage of firms is that high that the distribution of

Figure 11 — Stable and Unstable Equilibria



($\alpha = 0.1$, $\beta = 0.2$, $\varepsilon = 0.45$, $\mu = 0.25$, $\pi = 0.7$, $\sigma = 5$)

[18] See Rosenstein-Rodan (1943) and Murphy, Shleifer and Vishny (1989) for „big-push" theories. Last century California may provide an illustrative example. For a long time, the region could not establish an industrial base. Because demand was so low, it was cheaper to transport industrial goods from the east than producing these goods with a very low scale of production. The exploration of gold in the area induced a big push that suddenly created enough domestic demand that sustainable industrial production was possible. Cumulative processes led to further attraction of firms.

industries is illustrated by a point above the dotted line, such as in point C, industrial activities are sustainable. Then cumulative processes lead to a migration process of firms and workers till firms are equally distributed between the two regions. Below the dotted line and in the case of low trade costs firms always end up completely concentrated in one region.

4. Introducing Congestion Effects

As was shown in Figure 9 and 11 the economy ends up completely concentrated in one region if trade costs are at a low level; for high levels of trade costs, an equal distribution of industries results from an adjustment process. For intermediate trade costs, both equal distribution or complete concentration can result (Figure 10 and 11). However, the existing pattern of industry location in Europe or North America is very different, with fairly stable center-subcenter patterns. If the models are of any empirical relevance, then the level of trade costs between the most integrated regions like the US states or the regions in the European Union should be characterized as low. For low trade costs, however, the models predict a tendency towards complete concentration of firms in one region. If anything, the trend goes in exactly the opposite direction. Molle and Boeckhout (1995) describe converging patterns of industrial location over the Postwar period for Europe. Kim (1995) finds that regional specialization rose substantially in the US until the turn of the century – a period during which trade costs were much higher than today. He reports that specialization flattened out in the interwar years and fell enormously since the 1930s – a period during which trade costs fell substantially. Cheshir (1995) points out the great variety of population patterns and trends in the European Union. He shows that some centers are gaining and others loosing around a general distribution with centers and subcenters.

Thus, the prototype model can neither explain the observed existence of center-subcenter patterns, nor can it explain the current trend towards industry dispersion in Europe and the US. Traditional location theorists mentioned a number of centrifugal forces that would stop a trend towards agglomerations. Weber (1909) pointed out that higher land rents and wages might push industries out of the industrial core. Hoover (1948) added higher prices for raw materials in agglomerations. Isard (1956) mentioned higher land rents and a crowded transportation system as diseconomies of scale. Congestion effects in general have been modeled by Henderson (1975) and Kanemoto (1975) in the context an optimal city size.

Congestion effects can bind the cumulative forces towards agglomeration and prevent complete concentration due to negative externalities such as pollution, crime and bottlenecks in the infrastructure like traffic jams. That is, congestion effects can limit the attractiveness of being in the core in direct and indirect ways. They directly reduce the advantages for firms being in the core, if the costs of production increase through bottlenecks in the infrastructure or through pollution. Congestion directly reduces the advantages for workers being in the core, if workers value „soft" location factors. Then higher crime rates and pollution directly decrease the utility of the core as a residential location.

Besides these direct effects, congestion may have indirect effects if industry concentration leads to further rules and regulations that increase the costs of production. This can be the case if, for example, strong environmental pollution leads to public protest and hostility towards an expansion of industrial production, new airport runways, incinerating plants, power plants etc. Then, congestion effects could lead to political actions resulting in the imposition of regulations, which in turn raise the costs of production. Such regulations depend on the sensitivity of a society towards negative externalities of industry concentration. Societies that value environmental protection more, also regulate their industries more than others.

Congestion sensitivities might differ among regions for various reasons. One is the absolute level of pollution in a region. Another is that people in different regions have different rates of time preference. Future detrimental effects of pollution are considered more in some regions than in others. For instance, high income regions might value environmental protection and cleanness more than low income regions, which have more basic concerns like nutrition, health care or minimum social standards.

Congestion effects can also influence the model of section 2 in two ways. First, it can influence location utility directly, if people make their moving decisions on the base of real wages and soft location factors. In this case real wages would be adjusted by a term indicating the degree of pollution. Second, congestion can influence location utility indirectly by tighter regulations for industrial production. This leads to high production costs and lower real wages through the market process.[19]

I use the second approach for two reasons. First, it enables the analysis of congestion in a more specific way, because it allows a separate analysis of the cases where congestion influences fixed and variable costs. Second, I believe that it is empirically the more important case. I model the size of congestion effects in dependence of the number of firms in a region. In order to consider that nature and people are able to absorb a certain degree of pollution and congestion with few harm, I argue that congestion effects rise exponentially with the number of firms so that they barely matter for a low number of firms and get increasingly important for a large number of firms per region. I consider two cases:

[4.1] $$Z_{cj} = \alpha + \beta e^{\xi N_c} Q_{cj}$$, and

---

19 The first approach has been adopted by Asilis and Rivera-Batiz (1994). The second approach has been adopted by Brakman et al. (1993) and Junius et al. (1995).

[4.2] $$Z_{cj} = \alpha e^{\xi N_c} + \beta Q_{cj} \ ,$$

where ξ is the sensitivity of the society towards congestion effects. In equation [4.1], the input demand function [2.3] is replaced by a function where congestion effects raise variable costs. In equation [4.2], congestion effects raise fixed costs. In both cases, an increase of the number of firms in a region has an ambiguous effect on the costs of production. Centrifugal as well as centripetal forces are dependent on the number of firms per region. As before, a higher number of firms increases the number of goods produced in that region and reduces the regional price index for intermediate inputs. However, the negative externalities associated with the production of goods rise as well and raise variable or fixed costs.

Congestion Effects and Variable Costs

Including variable costs congestion effects changes some other equations of the model. Congestion increases the price of goods by the same proportion as the costs of production. This reflects that more inputs have to be used for the same output. Instead of equation [2.5], the following pricing conditions have to be adopted:

[4.3] $$\beta e^{\xi N_c} Q_{cj} = p_{cj} = (\sigma / \sigma - 1)\beta e^{\xi N_c} \ .$$

This directly leads to a different output per firm, and equation [2.6] has to be transformed to:

[4.4] $$Q_{cj} = (\sigma - 1)\alpha e^{-\xi N_c} \cdot \beta^{-1} \ .$$

Equation [4.4] shows that firms try to accommodate higher variable costs by a lower scale of production. This reduction of output per firms does not increase the number of firms, because it is not accompanied by a reduction of inputs used per firm. Thus, the number of firms can still be described by equation [2.11]. Marginal costs, however, increase directly with the size of the congestion effects. Equation [2.10] has to be replaced by:

[4.5] $$MC_c = w_c^{1-\mu} T_c^{\mu} \beta e^{\xi N_c} (1-\mu)^{\mu-1} \mu^{-\mu} .$$

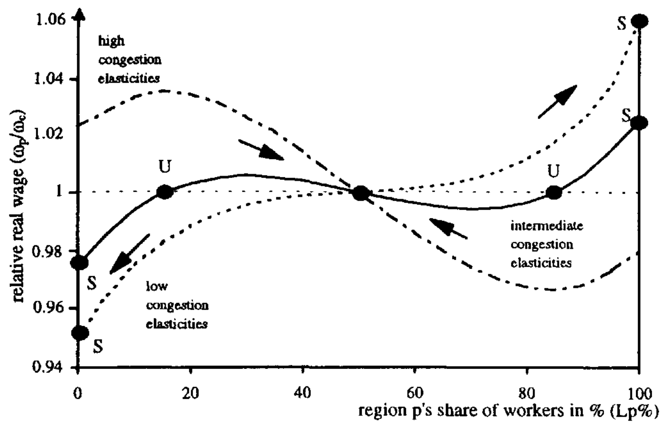
In turn, higher marginal costs raise the price index and lower the wage. Both can still be written as in equation [2.9] and [2.14], because MC and T_c appear as variables in the equations only.

Solutions to this model are illustrated in Figure 12. Figure 12 shows the relative real wage of the periphery for alternative industry distributions and alternative levels of variable costs congestion effects. Trade costs are set at a low level, so that linkage effects provide the larger region with an advantage if congestion is not considered (see Figure 9). For low congestion elasticities, the outcome of the model is not qualitatively distinct from Figure 9. The larger region always exhibits a higher real wage. For high congestion elasticities, the pattern is reverse. The smaller region always exhibits a higher wage. For intermediate congestion elasticities, an interesting pattern emerges. For very small regions (to the left from the first unstable equilibrium and to the right from the second unstable equilibrium), wages are lower than in the larger region. In this case, lower congestion costs cannot outweigh the disadvantages of fewer industry linkages. However, if the regions are more similar, congestion matters and makes production in the core more expensive than in the periphery.[20]

---

20 This suggests that for countries that are quite similar in their size lower congestion elasticities due to for instance fewer environmental concerns or lower industry density in one country could lead to a relocation of firms and workers in the slightly smaller country.

Figure 12 —Relative Wage for Alternative Industry Distributions (Variable Costs Congestions Effects)



($\alpha = 0.1$, $\beta = 0.4$, $\pi = 0.7$, $\mu = 0.2$, $\sigma = 4$, $\varepsilon = 0.4$, $\tau = 1.6$, $\xi_{low} = 0.17$, $\xi_{int} = 0.19$, $\xi_{high} = 0.23$)

Congestion Effects and Fixed Costs

Including fixed costs congestion effects also changes some of the equations of the model. Instead of the pricing equation [2.5] one gets:

[4.6] $$\beta + \alpha e^{N_c} Q_{cj} = p_{cj} = (\sigma / (\sigma - 1))\beta \ .$$

Here, congestion affects fixed costs, so average costs are higher than before for a given output. Under monopolistic competition, free entry of firms assures that firms produce output such that prices equal average costs. Since prices are also determined by the condition that marginal revenue equals marginal costs and since congestion does not directly affect marginal costs, prices (in input terms) remain the same. This means that firms have to produce a higher output to accommodate higher fixed costs such that average costs equal their level without congestion. Instead of equation [2.6], output per firm is then given by:

[4.7] $$Q_{cj} = (\sigma - 1)\alpha e^{\xi N_c} \cdot \beta^{-1} .$$

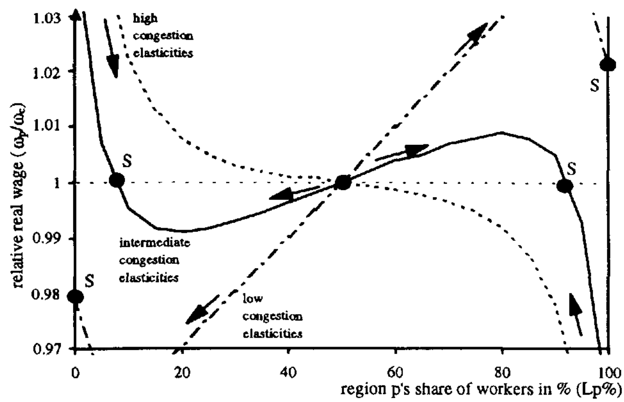
Because congestion is dependent on the number of firms per region, congestion effects raise fixed costs more in the larger than in the smaller region. This means that output per firm cannot be the same in both regions anymore. Equation [4.7] shows that the firms in the larger region produce with a larger scale of output than firms in the smaller region. A higher output per firm and a higher amount of inputs per output needed in the presence of congestion means that the fixed amount of inputs available in a region results in a smaller number of firms per region. Instead of equation [2.11], the number of firms is given by:

[4.10] $$N_c \cdot e^{\xi N_c} = \frac{L_c}{\alpha\sigma}\left[\frac{\mu}{1-\mu}\frac{w_c}{T_c}\right]^{\mu} .$$

Because congestion effects are lower in the smaller region, the number of firms reduces by relatively more in the larger than in the smaller region. Consequently, the regions are getting more similar in their economic variables that depend on the number of firms in a region like the price index, the marginal costs and the wages. These equations, however, can still be written as in equations [2.9], [2.10] and [2.14].

Solutions to this model are illustrated in Figure 13. Figure 13 shows the relative real wage of the periphery for alternative industry distributions and alternative levels of fixed costs congestion effects. Again trade costs are set to a low level so that linkage effects provide the larger region with an advantage, if congestion is not considered. Not surprisingly, for low congestion elasticities there is a similar pattern as without them (see Figure 9). Real wages of the peripheral region are lower than in the core. This can lead to complete concentration in the larger region. For a very high congestion elasticity, pollution abatement costs, regulation, crime and other congestion effects are so high that the larger region always exhibits lower wages than the smaller region. This leads to equal distribution of industries between regions.[21]

Figure 13 — Relative Wage for Alternative Industry Distributions (Fixed Costs Congestion Effects)



($\alpha = 0.1$, $\beta = 0.32$, $\pi = 0.7$, $\mu = 0.55$, $\sigma = 4$, $\varepsilon = 0.4$, $\tau = 1.5$, $\xi_{low} = 0.35$, $\xi_{int} = 4.25$, $\xi_{high} = 5$)

In between these two cases, there is another pattern of relative wages. For intermediate levels of congestion elasticities Figure 13 shows an S-shaped curve. In this case, there are three equilibria, two of which are stable. Complete concentration is no equilibrium anymore. In any long run equilibrium both regions produce industrial as well as agricultural products, although one region is larger than the other. Hence, for certain parameter constellations, the inclusion of congestion effects can explain a stable center-subcenter pattern of industry location as one finds

21 The empirical relevance of this case is questionable. Nevertheless, this argument has been widely used in policy debate. For instance in the 1992 presidential campaign Ross Perot warned American voters about the effects of NAFTA. He predicted that lower environmental standards in Mexico would lead to a strong movement of US jobs to Mexico unless wage levels converge.

it in most developed regions and free trade areas. Therefore, the prototype model extended by fixed cost congestion effects offers a theoretical possibility to explain existing patterns of industry agglomeration in the presence of centrifugal and centripetal forces.

5. The Impact of Land Rents and Commuting Costs

This section shall examine in how far other factors than fixed costs congestion effects can describe partly dispersed, partly concentrated patterns of industry location. One alternative factor, mentioned in traditional location theory, was the argument that land rents are higher in inner cities. This might push workers and industries to the suburbs or rural areas. Krugman and Livas (1992) incorporate this argument in their economic geography model. In their model, all production takes place at the central business district (CBD) and workers live along a straight line. The further they live from the CBD the longer they have to commute. Commuting in turn reduces possible work time, so that a worker who is living δ units from the CBD works only $(1-\delta)$ hours. His income is δw times lower than that of a worker who does not have to commute. However, a worker has to pay land rents that exactly offset the opportunity costs of commuting. Hence, real wages in this economy are given by:

$$\omega = \frac{w}{T^{\varepsilon}}(1-\delta) \; . \qquad [5.1]$$

Comparing two regions as in the prototype model of section 2, average commuting time and land rents are higher in the larger region. The smaller region has an advantage, as average travel distance to their CBD is shorter. Their workers save on commuting time, but face lower wages and a higher price index. The level of trade costs determines, which of the two effects prevails.

Brezis and Krugman (1993) argue along similar lines in a slightly different model. They analyze the spatial effects of technological change. When a new technology is developed and produced additionally to an existing one, „new technology" firms do not gain from being located in the CBD, if productivity is dependent on past experience in a technology. Thus, they do not set up production in the CBD, but in the periphery, where commuting costs and land rents are lower. If the new technology is superior to the old one, this leads to the decline of the old core and to the rise of a new core. Thus, new technologies can explain the decline and rise of cities and a reversal of a cumulative process towards industrial concentration in one city.

These approaches remain unconvincing as an explanation of large scale agglomerations, for two reasons. First, although land rents might be powerful argument in an urban context to explain industry location, it is less obvious that they are also relevant in a regional or a cross-country context. Second, one implication of the suggested models is that high land rents should push people with low income out of inner city districts, whereas workers with high opportunity costs of commuting, should live closer to their work places. Since reality is different, as can be seen in almost all US cities, there must be other disincentives of living in a densely populated area that outweigh commuting costs. Such disincentives seem to be higher crime rates in inner cities or higher pollution. These are exactly urban congestion effects. Therefore, I conclude that congestion effects as modeled in section 4 play a more decisive role than land rents as a centrifugal force, especially for the explanation of large scale agglomeration patterns.

6. The Impact of Self-fulfilling Expectations

Another reason why there may not be a tendency for complete concentration as implied by the prototype model of section 2 could be that with a different modeling of expectations, the dynamics of the model change and a different equilibrium is reached than the one suggested by the initial distribution of labor. It was assumed that workers and firms have static expectations with regard to wages and profits. That is, they believe that current wage differentials between the regions will persist in the future. If so, changes in regional employment are an increasing function of the wage gap:[22]

$$[6.1] \qquad -\dot{L}_p = \dot{L}_c = g(\omega_c - \omega_p), \quad g' > 0,$$

where dots over a variable denote changes in time. Whenever workers realize differences in regional real wages, they move to the region that pays higher wages. The existence of forward and backward linkages leads to external economies of scale. Thus, the wage gap between the two regions is an increasing function of their relative size:

$$[6.2] \qquad \omega_c - \omega_p = f(L_c - L_p), \quad f' > 0.$$

For simplicity, I consider the case of low trade costs only so that the core always pays higher wages than the periphery. Then f is a continous and strictly monotonic increasing function, and the initial distribution of industries is the only determinant for the dynamic process that leads to complete concentration of industries in the initially larger region.[23]

The simplicity of the adjustment process crucially depends on the neglect of moving costs and self-fulfilling expectations. With their inclusion the adjustment process is more difficult than above and the dynamics can work in a different way. If moving is costly, workers are concerned about current and future wage differentials between the regions. However, they can only observe current wages. Since they are not sure about the underlying model of the economy and its parameters, they do not know which region pays higher wages in the future. Nevertheless, workers have some idea about the economy and form expectations on the base of their

---

22 The dynamics then can be described by a Marshallian adjustment mechanism as in equation [6.1] and can be graphed as in Figure 10.

23 For high trade costs, the dynamics go the opposite way. This does not change the qualitative outcome of this section.

subjective view about it. This opens up the possibility of self-fulfilling expectations. If all people think that for some exogenous reason region p is going to be the most attractive location for industries, they move there. This strengthens market linkages in that region until it eventually becomes the superior location, despite an initial disadvantage. Thus, one of the reasons why cumulative processes might start or stop is just because people believe that they do so.[24]

For instance, political events or technological change can raise positive or negative expectations towards a region. An example might be the accession of Hong Kong to China that might increase uncertainty about its future prospects as a financial center. This could lead some people or firms to the believe that future wages and profits fall in Hong Kong. Such a development would strengthen the relative attractiveness of alternative locations such as Singapore. Hence, some migration would strengthen linkage effects in Singapore. A cumulative process would raise relative wages in Singapore and indeed lower them in Hong Kong, even if China had not changed its policy. Only self-fulfilling expectations would have led to Hong Kongs relative decline.

To find out when the initial distribution of industries and when self-fulfilling expectations play the decisive role for the dynamics of the economy, one has to formalize the above arguments regarding the calculus of workers.[25] Workers are assumed to be able to borrow and lend freely at interest rate r and to live infinitely. They maximize the present value of expected life-time wages, which at time x can be expressed as:

---

[24] If moving is not costly, but workers form expectations about future wages, the initial distribution of industries is irrelevant, because moving is costlessly reversible. Workers then move towards the region they believe pays higher wages and move back, if their expectations do not materialize or their expectations for wages in the next period favor another region.

[25] For seminal papers see Krugman (1991c) and Matsuyama (1991).

[6.3] $$H = \int_x^{\infty} \omega(t) \cdot e^{-rt} dt.$$

Workers will move to the other region, if the benefits of being there exceed the moving costs. In order to ensure that workers move smoothly, moving costs are assumed to rise with the square of workers moving per period. If $1/\gamma$ is a measure of these moving costs, γ indicates the speed of moving. Besides this technical speed of moving, the number of workers moving at a point of time ($\dot{L}_c$) depends on the value of working in the region of choice (q).

[6.4] $$\dot{L}_c = q\gamma.$$

At time x the present value of working in the region of choice depends on all current and future wage differentials and the discount factor (r):

[6.5] $$q(x) = \int_x^{\infty} \left(\omega_c(t) - \omega_p(t)\right) e^{-r(t-x)} dt \ .$$

Because moving is subject to moving costs, being in the region of choice is an asset and moving an investment decision. The equilibrium return per time period on the asset of being in the region of choice should equal the difference of wages plus the capital gain on this asset:

[6.6] $$rq = \left(\omega_c - \omega_p\right) + \dot{q} \ .$$

Without any loss in the economic content, I respecify equation [6.2] by

[6.7] $$\omega_c - \omega_p = \theta L_c, \quad \theta > 0 \text{ and } \omega_c = \omega_p = 1 \text{ if } L_c = L_p,$$

where θ indicates the degree of external economies of scale. As a consequence [6.6] and [6.7] can be rearranged to:

[6.8] $$\dot{q} = rq - \theta L_c \ .$$

Equation [6.4] and [6.8] constitute a dynamic equation system in the (q, L_c) space. [26] The qualitative solution of the dynamic system can be graphed in a phase diagram (see Figure 14).[27] The objective of a phase diagram is to translate the dynamics implied by two differential equations into a system of arrows that describe the qualitative behavior of the economy over time.

The horizontal axis of Figure 14 shows region c's share of workers. The vertical axis shows the value of being in region c rather than in region p. Each point in the phase diagram represents a position of q and L_c at a given point of time. To construct the phase diagram, I first determine the $\dot{L}_c = 0$ and $\dot{q} = 0$ schedules. The $\dot{L}_c = 0$ schedule shows all combinations of L_c and q, where L_c does not change over time, i.e. where no migration between the two regions takes place. From inspection of equation [6.4] one can see that this is only the case for

---

[26] Fukao and Benabou (1993) criticize the formulation of q as in equation [6.5] in their comment on Krugman (1991c). Instead of integrating over an infinite life span, one should integrate until a finite point of time, at which concentration of all workers in one region is reached. Then the last movers have an incentive to avoid moving costs by waiting until everybody else has moved to the region of choice. The interpretation of equation [6.4] then is that the moving costs equal what workers are willing to pay for moving at a point t rather than at the finite point of time, when moving becomes costless. This would lead to different terminal conditions and equilibrium paths of the economy towards the full concentration points. This approach is not adopted her, because it invalidates the arbitrage condition [6.8], besides the different interpretation of equation [6.4]. This means that in Fukao and Benabou's formulation of the model, there is no value associated with being in the fully concentrated region of choice, so that q=0 for L_c = 100%, which I consider to be a problematic economic implication.

[27] See Chiang (1984) or the mathematical appendix of Barro and Sala-i-Martin (1995) for a comprehensive explanation of the working of phase diagrams.

$q = 0$. If the value associated with being in another region is zero, no migration takes place. The $\dot{L}_c = 0$ schedule, therefore, equals the horizontal axis.

The $\dot{q} = 0$ schedule shows all combinations of L_c and q where q does not change over time, i.e. where the value of the asset q remains constant. From inspection of equation [6.8] one can see that this is only the case if the wage gap between the regions equals the interest rate times the asset q. Since the wage gap equals zero if the regions are of equal size, $L_c = 50\%$; $q = 0$ can be determined as one point of the schedule. Solving equation [6.8] for q at $\dot{q} = 0$ gives $q = (\theta/r)\ \bar{L}_c$. Then the slope of the schedule is given by $dq / dL_c = \theta / r$, and the schedule can be drawn as in Figure 14.

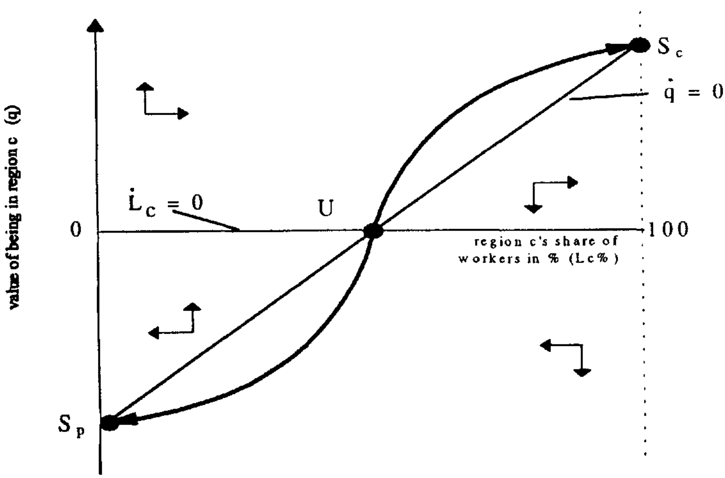
The $\dot{q} = 0$ and the $\dot{L}_c = 0$ schedules intersect at the point ($L_c = 50\%$; $q = 0$). This is an intertemporal equilibrium point. At any other point either q, L_c or both change over time. The direction of these changes can be calculated and expressed by the laws of motion indicated by the little arrows.

Horizontal arrows indicate how L_c changes for a given q. An arrow pointing to the right means that L_c grows, i.e. workers move from region p to region c. An arrow pointing to the left means that L_c falls, i.e. workers move to region p. Because workers move to region c if q is positive, arrows are pointing to the right above the $\dot{L}_c = 0$ schedule. Because workers leave region c if q is negative, arrows are pointing to the left below the $\dot{L}_c = 0$ schedule.

Vertical arrows indicate how q changes for a given L_c. An arrow pointing upwards means that q increases, i.e. the asset of being in region c rises over time. An arrow pointing downwards means that q falls, i.e. the asset of being in region c falls over time. The direction of the arrows can be determined by the derivative of the $\dot{q}$ schedule with respect to q, i.e $d\dot{q} / dq = r > 0$. This means that $\dot{q}$ steadily increases with a higher q. If one moves continually from the bottom to the top of the phase diagram crossing the $\dot{q} = 0$ schedule $\dot{q}$ must pass through three stages in the order negative, zero and positive. This means that vertical arrows point downwards below the $\dot{q} = 0$ line and point upwards above it. The economic interpretation is that above the $\dot{q} = 0$ schedule, the wage differential between the regions is lower than the interest rate times the asset of being in region c. For this to hold the expected gain on the asset of being in region c must be rising (see equation [6.6]).

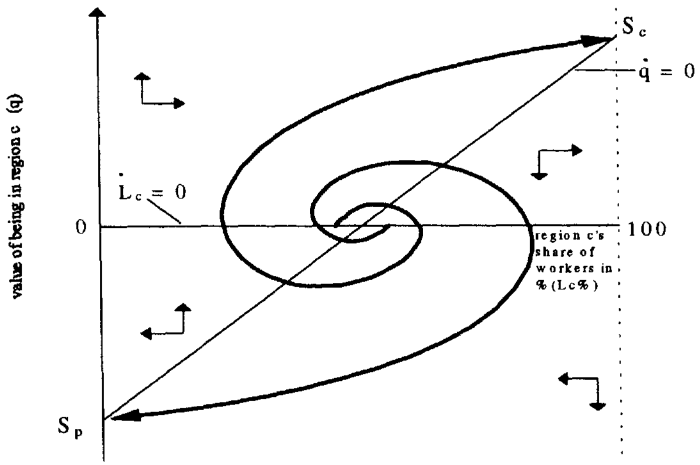
The dynamics of the system consistent with the laws of motion can take the shape of an S, as in Figure 14. For values of L_c < 50% the dynamics lead to a stable equilibrium (S_p) with zero % workers in region c. For values of L_c > 50% the dynamics lead to a stable equilibrium (S_c) with full concentration of all workes in region c. Thus, the initial distribution of industries determines, which equilibrium is reached.

Figure 14 — Initial Distributions Determine the Adjustment



However, two interlocking spirals as in Figure 15 are also consistent with the laws of motion. In this case the dynamics can lead to complete concentration in either region from a larger number of initial distribution of industries. This means that the economy can end up fully concentrated in the region that initially was smaller and paid lower wages. Consequently, self-fulfilling expectations instead of initial industry distribution can determine, which equilibrium is reached.

Figure 15 — Expectations Can Determine the Adjustment



In order to analyze when the initial distribution only and when also expectations can determine the long run equilibrium, I determine the local stability of the system at the intersection point of the $\dot{q} = 0$ and the $\dot{L}_c = 0$ schedules.[28] Thus, one has to find solutions for:

[6.9] $$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} \dot{q} \\ \dot{L}_c \end{bmatrix} + \begin{bmatrix} -r & \theta \\ -\gamma & 0 \end{bmatrix} * \begin{bmatrix} q \\ L_c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

$q(t) = me^{bt}$ and $L_c(t) = ne^{bt}$ are possible general solutions for such a system.[29] From that it follows that $\dot{q} = bme^{bt}$ and $\dot{L} = bne^{bt}$. Then, [6.9] can be transformed to:

[6.10] $$\begin{bmatrix} b & 0 \\ 0 & b \end{bmatrix} * \begin{bmatrix} m \\ n \end{bmatrix} + \begin{bmatrix} -r & \theta \\ -\gamma & 0 \end{bmatrix} * \begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which equals:

[6.11] $$\begin{bmatrix} b-r & \theta \\ -\gamma & b \end{bmatrix} * \begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

In order to find a non-trivial solution for m and n the determinant of the first matrix has to equal zero:

[6.12] $$b^2 - br + \theta\gamma = 0.$$

From that one gets the characteristic roots of the system that equal:

[6.13] $$\frac{r +/- \sqrt{r^2 - 4\theta\gamma}}{2}$$

---

[28] A local stability analysis only determines, which equilibrium is reached from a specific point. From any other point a different equilibrium could result. However, Fukao and Benabou (1993) show that the global behavior of this system equals its local behavior and that the economy indeed hits a boundary at $L_c = 0$ or $L_c = 100\%$ in finite time. Therefore, I need to analyze the local stability only.

[29] See Barro and Sala-i-Martin (1995: 468–471).

The characteristic roots indicate, what form the dynamics of the system take, i.e. if they directly lead to one of the two stable equilibria at S_p and S_c, if they lead to one of the equilibria after fluctuating around the initial equilibrium point (at the intersection of the $\dot{q} = 0$ and $\dot{L}_C = 0$ schedules) or if they never reach an equilibrium. In principle six cases can be distinguished. If the roots are real numbers and positive, the initial equilibrium point is unstable and the dynamics lead directly to concentration in one of the two full concentration equilibria. If the roots are real numbers and negative, the initial equilibrium point is stable and the dynamics lead always to equal distribution of industries between the two regions. If the roots are real numbers but of opposite sign, the system is saddle path stable. If the roots are complex with negative real parts, the economy converges to the initial equilibrium point in an oscillating manner and equal industry distribution results. If the roots are complex with positive real parts, the economy diverges from the initial equilibrium point in an oscillating manner and concentration in one of the regions results. If the roots are complex with zero real parts, the system does not approach any equilibrium, but infinitely circles around the initial equilibrium point.

In the present context, two cases matter. The real parts, which equal r here, are positive in any case, because only positive interest rates were considered. Therefore the system has two positive roots for $r^2 > 4\theta\gamma$, and two complex roots for $r^2 < 4\theta\gamma$.

With two positive roots, no fluctuation of the system is possible. It approaches directly its long-run equilibria. Then the adjustment process can be described by Figure 14, where the economy always ends up in the initially larger region and expectations play no role.

With two complex roots, the economy evolves in an oscillating manner. The positive real parts assure that the system fluctuates in widening oscillations. This means that the economy diverges from the initial equilibrium and ends up concentrated in one region. Due to the fluctuations around the initial equilibrium point, the system can end up concentrated in the initially smaller region. Then the adjustment process can be described by Figure 15. If the economy ends up concentrated in the initially smaller region, self-fulfilling expectations must have played a role. Hence, expectations matter if $r^2 < 4\theta\gamma$.

The role of self-fulfilling expectations thus depends on the relationship of r, θ and γ. Self-fulfilling expectations are possible if the interest rate (r) is low, economies of scale (θ) are high and the adjustment process (γ) is fast. This is because for low interest rates, workers discount future wages relatively little. Then, expected future wages matter quite a lot. If economies of scale are large, the interdependence of workers' expectations and decisions is relatively strong. Finally, if the speed of adjustment is fast, expectations matter relatively much, because the initial distribution of workers is less decisive for the determination of future periods' wage levels.[30]

This section has shown that self-fulfilling expectations can stop cumulative processes and explain a movement of industries and workers out of the initially advantaged region. Self-fulfilling expectations can play a role if the rate of time preference and moving costs are low and external economies of scale are high. If so cumulative causation will not necessarily lead the economy to concentration in the initially larger region. It can approach any equilibrium. In how far self-fulfilling expectations actually play a role in explaining the observed trends of industry distribution in the US and Europe is questionable. I would assume that self-fulfilling

[30] This might explain why labor moves much faster according to current wage differentials in the US than in Europe. Implicit moving costs are lower in the US than in Europe, as language, cultural and legislative barriers are lower. Moving is therefore more reversible than in Europe. The rate of time preference and the size external economies of scale may also play a role, but differences in these parameters between the US and Europe are less obvious.

expectations play a more decisive role in the determination of lower scale agglomerations, which means at an urban rather than at a regional level. In any case, self-fulfilling expectations fail to explain a center-subcenter pattern of industry location.

7. Summary

In this paper, I set up an economic geography model to explain industrial location patterns for two trading regions. The model suggests that forward and backward linkages lead to external economies of scale. The demand of a dispersed immobile part of the population works against this centripetal force. For high trade costs, a cumulative process leads to equal distribution of industries. For low trade costs, complete concentration in the initially larger region is the result.

This finding is not supported by empirical evidence. At least in the US and in Europe, center-subcenter patterns of industry distribution continue to exist, and a clear trend towards more pronounced agglomeration and concentration patterns cannot be observed. Therefore, I suggested the inclusion of further centrifugal forces. Congestion effects, land-rents and self-fulfilling expectations are possible reasons for cumulative processes to peter out or to reverse. Out of these congestion effects seem to be the most relevant in an international or interregional context. Therefore, I include congestion effects in a prototype economic geography model. The augmented model can explain the existence of stable center-subcenter patterns of industry location.

The dynamics of the model depend on a number of factors. Congestion effects and economies of scale are the crucial parameters that determine the equilibrium pattern of industry. Further research should focus on the empirical relevance of scale economies and congestion effects in order to get a better understanding of real world agglomeration processes.

Appendix

List of Variables and Parameters Used in the Model

Indices:

c, p	-	index for regions (core and periphery)
$j = 1....(N_c+N_p)$	-	index for industrial goods
M, A	-	index for sectors (industry and agriculture)

Parameters and variables:

α	-	fixed costs parameter (>0)
β	-	variable costs parameter (>0)
γ	-	speed of the adjustment process (inverse index of moving costs)
ε	-	share of actual income spent on industrial goods
λ	-	Lagrangemultiplier
μ	-	share of intermediate goods in the inputs of a good
ξ	-	congestion parameter
π	-	share of workers in overall population ($0 < \pi < 1$)
σ	-	elasticity of substitution (inverse index of internal scale economies), ($\sigma > 1$)
θ	-	indicator of external scale economies ($\theta > 0$)
τ	-	amount of goods that have to be shipped so that one good arrives (iceberg trade costs) (τ>1)
ω	-	real wage
C	-	commodity group
c	-	single product variety
e	-	elasticity of demand

H	-	present value of life time wages
k	-	parameter (>0)
K	-	total costs of production
L	-	industrial workers
M	-	aggregate of all intermediate goods
MC	-	marginal costs
m	-	single intermediate good
N	-	number of potential firms and product varieties
p	-	price of an industrial good expressed in input terms
q	-	value of the asset of having a job in region c rather than in region
Q	-	output
r	-	interest rate
S	-	sales of all firms of a region
t	-	point of time
T	-	price index
U	-	utility
w	-	wage
W	-	wage sum in the industrial sector
x	-	initial point of time
X	-	expenditure on industrial goods
Y	-	income
Z	-	inputs

1. Derivation of Equation [2.10], the Marginal Costs

Solving [2.8] for L_{cj} yields:

[A.1] $$L_{cj} = \frac{1-\mu}{\mu}\frac{M_j T_c}{w_c}$$

Inserting [A.1] and [2.4] into [2.3] yields:

[A.2] $$\left(\frac{1-\mu}{\mu}\frac{T_c}{w_c}\right)^{1-\mu} M_j = \alpha + \beta Q_{cj}.$$

After rearranging one gets:

[A.3] $$M_j = (\alpha + \beta Q_{cj})\, T_c^{\mu-1} w_c^{\mu-1} \left(\frac{1-\mu}{\mu}\right)^{\mu-1}.$$

The total costs of production of a firm equals the sum of cash outlays for labor and intermediate goods:

[A.4] $$K_{cj} = w_c L_{cj} + T_c M_j.$$

Inserting [A.1] into the above costs function gives:

[A.5] $K_{cj} = M_j T_c (1-\mu)/\mu + M_j T_c$, which equals:

[A.6] $$K_{cj} = M_j T_c/\mu .$$

Inserting in [A.3] for M_j in [A.6] gives:

[A.7] $$K_{cj} = \mu^{-1} T_c\, T_c^{\mu-1} (1-\mu)^{\mu-1} \mu^{1-\mu} w_c^{1-\mu} (\alpha + \beta Q_{cj}).$$

After rearranging one gets the costs of production as a function of output and factor prices:

[A.8] $$K_{cj}(Q_{cj}, T_c, w_c) = T_c^{\mu} w_c^{1-\mu} (1-\mu)^{\mu-1} \mu^{-\mu} (\alpha + \beta Q_{cj}).$$

From this one can derive the marginal costs:

[A.9] $$MC_c = \frac{dK_{cj}}{dQ_{cj}} = T_c^{\mu} w_c^{1-\mu} (1-\mu)^{\mu-1} \mu^{-\mu} \beta.$$

2. Derivation of Equation [2.12], the Number of Firms in a Region

Since all firms employ the same number of workers, the number of firms in a region can be expressed by:

[A.10] $$N_c = L_c/L_{cj} .$$

Inserting [2.6] and [2.4] into [2.3] gives:

[A.11] $$L_{cj}^{1-\mu}M_j^{\mu} = \sigma\alpha$$

Substituting M_j in [A.11] by the use of [2.8] yields to:

[A.12] $$L_{cj}^{1-\mu}\left(\frac{\mu}{1-\mu}\frac{w_c}{T_c}L_{cj}\right)^{\mu} = \sigma\alpha.$$

After substituting for L_{cj} by the use of [A.10] one gets:

[A.13] $$\sigma\alpha = \left(\frac{\mu}{1-\mu}\frac{w_c}{T_c}\right)^{\mu}\frac{L_c}{N_c}.$$

After rearranging the number of firms in a region results as:

[A.14] $$N_c = \frac{L_c}{\sigma\alpha}\left[\frac{\mu}{1-\mu}\frac{w_c}{T_c}\right]^{\mu}.$$

3. Derivation of [2.15], the Wage Rate

The expenditure of residents of region c on industrial goods from both regions is:

[A.15] $$X_c = \sum_{j=1}^{N_c+N_p} c_{jc}p_{jc}.$$

Any good c_c produced and sold in region c will be bought according to:

[A.16] $$\frac{c_{jc}}{c_c} = \left(\frac{p_c}{p_{jc}}\right)^{\sigma},$$

where c_{jc} is any other good from any region sold in region c for the price p_{jc}.

Multiplying by c_c and $p_c^{1-\sigma}$ one gets:

[A.17] $$p_c^{1-\sigma}c_{jc} = p_c c_c p_{jc}^{-\sigma}.$$

Multiplying by p_{jc} and taking the sum over j yields:

[A.18] $$p_c^{1-\sigma}\sum_{j=1}^{N_c+N_p} c_{jc}p_{jc} = p_c c_c \sum_{j=1}^{N_c+N_p} p_{jc}^{1-\sigma}.$$

Dividing by $\Sigma p_{jc}^{1-\sigma}$ gives the expenditure of c residents on good c_c:

[A.19] $$\frac{p_c^{1-\sigma}\Sigma c_{jc}p_{jc}}{\Sigma p_{jc}^{1-\sigma}} = p_c c_c.$$

Inserting [A.15] for $\Sigma c_{jc}p_{jc}$ gives:

[A.20] $$p_c c_c = \frac{p_c^{1-\sigma}X_c}{\Sigma p_{jc}^{1-\sigma}}.$$

In region c there are N_c such goods competing with N_c+N_p goods from both regions. Therefore, multiplying [A.20] by N_c gives total sales of region c's goods in its home market:

[A.21] $$S_{cc} = \frac{N_c p_c^{1-\sigma}X_c}{\sum_{j=1}^{N_c+N_p} p_{jc}^{1-\sigma}}.$$

Because prices equal marginal costs plus mark-up and transport costs, the price of a good produced and sold in c is $MC_c\sigma/\sigma-1$. The price of a good produced in p and sold in c is $\tau MC_p\sigma/(\sigma-1)$ Substituting for prices in [A.21] one gets:

[A.22] $$S_{cc} = \frac{N_c MC_c^{1-\sigma}X_c}{N_c MC_c^{1-\sigma} + N_p(\tau MC_p)^{1-\sigma}}.$$

Sales of goods of region c in region p are equivalently:

[A.23] $$S_{cp} = \frac{N_c(\tau MC_c)^{1-\sigma}X_p}{N_c(\tau MC_c)^{1-\sigma} + N_p(MC_p)^{1-\sigma}}.$$

Because the denominator of [A.22] equals $T_c^{1-\sigma}(N_c+N_p)(\sigma/\sigma-1)^{\sigma-1}$, one can use the definition of the price index from [2.9] and transform [A.22] and [A.23] to:

[A.24] $$S_{cc} = \frac{N_c}{N_c+N_p}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}X_c\frac{MC_c^{1-\sigma}}{T_c^{1-\sigma}},$$

[A.25] $$S_{cp} = \frac{N_c}{N_c + N_p}\left(\frac{\sigma}{\sigma - 1}\right)^{1-\sigma} X_p \frac{\tau MC_c^{1-\sigma}}{T_p^{1-\sigma}}.$$

Adding S_{cc} and S_{cp} yields the total sales of region c:

[A.26] $$S_c = \frac{N_c}{N_c + N_p}\left(\frac{\sigma}{\sigma - 1}\right)^{1-\sigma} (X_c[\frac{T_c}{MC_c}]^{\sigma-1} + X_p[(\frac{T_p}{\tau MC_c}]^{\sigma-1}).$$

Total revenue times the share of labor in production equals the sum of wages:

[A.27] $$(1 - \mu)S_c = w_c L_c \ .$$

[A.26] and [A.27] give the wage rate in region c.

[A.28] $$w_c = \frac{(1 - \mu)N_c}{(N_c + N_p)L_c}\left(\frac{\sigma}{\sigma - 1}\right)^{1-\sigma} [X_c(\frac{T_c}{MC_c})^{\sigma-1} + X_p(\frac{T_p}{\tau MC_c})^{\sigma-1}].$$

References

Amiti, M. (1995). Economic Integration, Industrial Structure and Trade Patterns. Center of Economic Performance, Discussion Paper 225.

Asilis, C., and L. Rivera-Batiz (1994). Geography, Trade Patterns, and Economic Policy. IMF Working Papers 94.16.

d' Aspremont, C., R.D.S. Ferreira, and L.-A. Gerard-Varet (1996). On the Dixit-Stiglitz Model of Monopolistic Competition. *American Economic Review* 86 (3): 623–629.

Barro, R., and X. Sala-i-Martin (1995). *Economic Growth.* McGraw-Hill.

Brakman, S., H. Garretsen, R. Gigengack, C. van Marrewijk, and R. Wagenvoort (1993). Congestion and Industrial Location. Research Memorandum 554, University of Groningen, Institute of Economic Research.

Brezis, E., and P. Krugman (1993). Technology and the life Cycle of Cities. NBER Working Paper 4561.

Cheshire, P. (1995). A New Phase of Urban Development in Western Europe? The Evidence for the 1980's. *Urban Studies* 32 (7): 1045–1063.

Chiang, A.C. (1984). *Fundamental Methods of Mathematical Economics.* McGraw-Hill, New York, 3rd edition.

Dixit, A., and J. Stiglitz (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review* 67: 297–308.

— (1993). Monopolistic Competition and Optimum Product Diversity: Reply. *American Economic Review* 83 (1): 302–304.

Ethier, W.J. (1982). National and International Returns to Scale in Modern Theory of International Trade. *American Economic Review* 72: 810–822.

Fujita, M., and P. Krugman (1995). When is the Economy Monocentric?: von Thünen and Chamberlin Unified. *Regional Science and Urban Economics* 25: 505–528.

Fujita, M., P. Krugman, and T. Mori (1995). On the Evolution of Hierachical Urban Systems. Kyoto Institute of Economic Research, Discussion Paper 419.

Fukao, K., and R. Benabou (1993). History Versus Expectations: A Comment. *Quarterly Journal of Economics* 108: 535–542.

Helpman, E., and P. Krugman (1985). *Market Structure and Foreign Trade*. 2nd ed. MIT Press.

Henderson, J. V. (1975). Congestion and Optimal City Size. *Journal of Urban Economics* 2: 48–62.

Hirschman, A.O. (1958). *The Strategy of Development*. Yale University Press, New Haven.

Hoover, E.M. (1948). *The Location of Economic Activity*. New York, McGraw-Hill.

Isard, W. (1956). *Location and Space Economy*. Cambridge, MIT Press.

Junius, K., E. Nijsse, E., and D. Urban (1995). Congestion and the Location of Vertically Linked Industries. Kiel Advanced Studies Working Paper 280. Institute of World Economics, Kiel.

Kanemoto, Y. (1975). Congestion and Cost Benefit Analysis in Cities. *Journal of Urban Economics* 2: 246–264.

Kim, S. (1995). Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860–1987. *Quarterly Journal of Economics* 110: 881–908.

Krugman, P. (1991a). *Geography and Trade*. MIT Press, Cambridge.

— (1991b). Increasing Returns and Economic Geography. *Journal of Political Economy* 99: 483–499.

— (1991c). History Versus Expectations. *Quarterly Journal of Economics* 106: 651–667.

— (1992). A Dynamic Spatial Model. NBER Working Paper 4219.

— (1993). On the Number and Location of Cities. *European Economic Review* 37: 293–298.

— (1994). Fluctuation, Instability and Agglomeration. NBER Working Paper 4616.

Krugman, P., and R. Livas (1992). Trade Policy and the Third World Metropolis. NBER Working Paper 4238.

Krugman, P., and A. Venables (1990). Integration and the Competitiveness of Peripheral Industry. In C. Bliss et al.(ed.), *Unity with diversity*. Cambridge Univ. Press.

— (1993). Integration, Specialization and Adjustment. CEPR Working Paper 886.

— (1995a). Globalization and the Inequality of Nations. *Quarterly Journal of Economics* 110 (4): 560–587.

— (1995b). The Seamless World: A Spatial Model of International Specialization. CEPR Diskussion Paper 1230.

Markusen, J.R. (1989). Trade in Producer Services and in Other Specialized Intermediate Inputs. *American Economic Review* 79 (1): 85–95.

Matsuyama, K. (1991). Increasing Returns, Industrialization, and Indeterminancy of Equilibrium. *Quarterly Journal of Economics* 106: 616–650.

Molle, W., and S. Boeckhout (1995). Economic Disparity under Conditions of Integration – A Long Term View of the European Case. *Papers in Regional Science: The Journal of the RSAI* 74 (2): 105–123.

Murphy, K., A. Shleifer, and R. Vishny (1989). Industrialization and the Big Push. *Journal of Political Economy* 97: 1003–1026.

Rosenstein-Rodan, P. (1943). Problems of Industrialisation of Eastern and South-eastern Europe. *Economic Journal* 53: 202–211.

Samuelson, P. (1954). The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments. *Economic Journal* 64: 264–289.

Venables, A. (1996). Equilibrium Locations of Vertically Linked Industries. *International Economic Review* 37 (2): 341–360.

von Thünen, J. (1825). *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie.* Hamburg, Perthes, English translation: Oxford Pergamon Press 1966.

Weber, A. (1909). *Standort der Industrien.* Mohr, Tübingen.

Yang, X., and B. Heijdra (1993). Monopolistic Competition and Optimum Product Diversity: Comment. *American Economic Review* 83 (1): 295–301.